

Constraints on Single-Field Inflation

David Pirtskhalava^a, Luca Santoni^{b,c}, Enrico Trincherini^{b,c}

^a*Institut de Théorie des Phénomènes Physiques
EPFL Lausanne, Switzerland*

^b*Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126, Pisa, Italy*

^c*INFN – Sezione di Pisa, 56200, Pisa, Italy*

Abstract

Many alternatives to canonical slow-roll inflation have been proposed over the years, one of the main motivations being to have a model, capable of generating observable values of non-Gaussianity. In this work, we (re-)explore the physical implications of a great majority of such models within a single, effective field theory framework (including novel models with large non-Gaussianity discussed for the first time below.) The constraints we apply—both theoretical and experimental—are found to be rather robust, determined to a great extent by just three parameters: the coefficients of the quadratic EFT operators $(\delta N)^2$ and $\delta N \delta E$, and the slow-roll parameter ε . This allows to significantly limit the majority of single-field alternatives to canonical slow-roll inflation. While the existing data still leaves some room for most of the considered models, the situation would change dramatically if the current upper limit on the tensor-to-scalar ratio decreased down to $r < 10^{-2}$. Apart from inflationary models driven by plateau-like potentials, the single-field model that would have a chance of surviving this bound is the recently proposed slow-roll inflation with weakly-broken galileon symmetry. In contrast to *canonical* slow-roll inflation, the latter model can support $r < 10^{-2}$ even if driven by a convex potential, as well as generate observable values for the amplitude of non-Gaussianity.

1 Introduction and summary

In the last decades cosmology has seen a remarkable transformation into a precision science. The ongoing experimental program aims at constraining the properties of the universe to an unprecedented accuracy, the ultimate goal being to shed light on the precise physics governing its dynamics.

Inflation provides perhaps the most compelling picture of the universe at the early stages of its history. Although extremely successful as a paradigm, it is fair to say that the details of the microscopic physics behind it are still far from clear: there are many models, and a certain fraction of these is capable of reproducing the current observational data quite well. While awaiting further experimental input, it is thus important to look for ways of prioritizing the existing list of inflationary scenarios.

A particularly convenient framework for carrying out model-independent analyses of single field inflation has been suggested in [1, 2], and goes under the name of the *effective field theory of inflation* (EFTI) (see Ref. [3] for a complementary approach to the problem.) The relevant effective theory is formulated in the *unitary gauge*, in which the inflation perturbations are frozen (or, in other words, ‘eaten’ by the metric), $\phi(x, t) = \phi_0(t)$. One then writes down all possible operators in derivative expansion, consistent with the unbroken symmetries; for single-field inflation, the latter are arbitrary reparametrizations of spatial coordinates, $x^i \rightarrow x^i + \xi^i(x, t)$, and the most general unitary-gauge action reads [2]

$$S = \int d^4x \sqrt{-g} \left[\frac{M_P^2}{2} \left({}^{(3)}R + \frac{E_{ij}E^{ij} - E^2}{N^2} \right) - \frac{M_P^2 \dot{H}}{N^2} - M_P^2 (3H^2 + \dot{H}) + \right. \\ \left. + \frac{M_2^4}{2} (\delta N)^2 + M_3^4 (\delta N)^3 - \hat{M}_1^3 \delta N \delta E + \hat{M}_2^3 (\delta N)^2 \delta E + \dots \right], \quad (1)$$

where by the ellipsis we have denoted terms of higher order in field perturbations¹ and/or spacetime derivatives. The building blocks of the relevant effective field theory are quantities covariant under the three-dimensional diffeomorphisms, such as the lapse variable N , the intrinsic and extrinsic curvatures of equal-time hypersurfaces ${}^{(3)}R$ and $K_{ij} = E_{ij}/N$, etc., see Appendix A for a detailed account. While the first line of (1) is fixed by the background equations of motion, the coefficients of the operators in the second line are unconstrained and parametrize all possible single-field models on a quasi de-Sitter space with a Hubble rate H [2].

The free parameters of the effective theory (1) depend on the underlying model of inflation. For canonical slow-roll inflation, for example, they all vanish. More generally, in the spirit of EFT, one expects that the physical observables are predominantly determined by operators with the least number of space/time derivatives. In particular, if the unitary-gauge action (1) stems from an effective field theory of the inflaton with a single characteristic cutoff scale Λ and with no unnatural parameters, one can show that the leading deviations from slow-roll inflation are determined by the coefficients M_2^4 , M_3^4 , etc. of the operators with no derivatives. The rest

¹By ‘ δ ’ we denote the perturbation of the relevant quantity over its background value. In what follows, we will be interested in the effective action up to the 3rd order in the field perturbations.

of the operators give rise to effects further suppressed by powers of the small ratio Λ/M_{Pl} , and the derivative expansion applies in the unitary gauge in a straightforward way.

The effective field theories of the type (1) with non-zero coefficients M_2^4 and M_3^4 arise from models such as k -inflation [4] and DBI inflation [5, 6] (strictly speaking, it is only the latter model that can be considered a well-defined EFT in terms of the original inflaton field.) The experimental constraints on DBI-like theories, described by the action (1) with $\hat{M}_1^3 = \hat{M}_2^3 = 0$ are well-known [7, 8]. A particularly interesting region of the parameter space is the one corresponding to strongly subluminal scalar perturbations, $M_2^4 \gg M_{\text{Pl}}^2 |\dot{H}|$, in which case there is the possibility to generate sizeable non-Gaussianity within the regime of validity of the low-energy effective theory. The existing experimental data still leaves room for detecting interesting deviations from canonical slow-roll inflation within the DBI framework, and indeed, any possible hints of (equilateral) non-Gaussianity would make a strong case for these models.

In this paper we wish to explore the status of a more general effective theory of inflation, characterized by non-zero \hat{M}_1^3 and \hat{M}_2^3 . At first sight, the above remarks concerning the derivative expansion imply that large effects from the corresponding operators should not be anticipated: they are higher-derivative, and are therefore expected to play a minor role. There are however cases where the effects associated with these operators can, and do, become large. This happens in theories characterized by *weakly broken* [9] *invariance under galileon transformations* [10], which in the unitary gauge take on the following form

$$t \rightarrow t + b_\mu x^\mu . \quad (2)$$

Whenever there exists an (approximate) invariance under (2), the magnitudes of the EFT coefficients in the action (1) are governed by the following *radiatively stable* hierarchy [9]

$$|M_2^4| \sim |M_3^4| \sim |\hat{M}_1^3 H| \sim |\hat{M}_2^3 H| . \quad (3)$$

It is then straightforward to show that all of the operators in (1) play an *equally important role* for the dynamics of perturbations. Furthermore, even for the simplest theories with weakly broken galileon (WBG) symmetry, the coefficients in (3) are generically independent of each other.

In order to proceed, one needs a guideline concerning magnitudes of the above EFT coefficients that can arise from a sensible inflationary model. In fact, there are good reasons to believe that the coefficients in (3) can not parametrically exceed $M_{\text{Pl}}^2 H^2$ in magnitude if the inflationary background under consideration is to be insensitive to loop corrections ² [9]. This motivates the following parametrization of the unitary-gauge effective theory (1)

$$\alpha = \frac{\hat{M}_1^3}{2M_{\text{Pl}}^2 H} , \quad \beta = \frac{M_2^4}{2M_{\text{Pl}}^2 H^2} , \quad \gamma = \frac{\hat{M}_2^3}{M_{\text{Pl}}^2 H} , \quad \delta = \frac{M_3^4}{M_{\text{Pl}}^2 H^2} . \quad (4)$$

The values of the above dimensionless coefficients will encode the deviations of a given model from canonical slow-roll inflation, as well as distinguish between various single-field models.

²The theories with WBG symmetry can saturate this upper bound within a well-defined low-energy EFT.

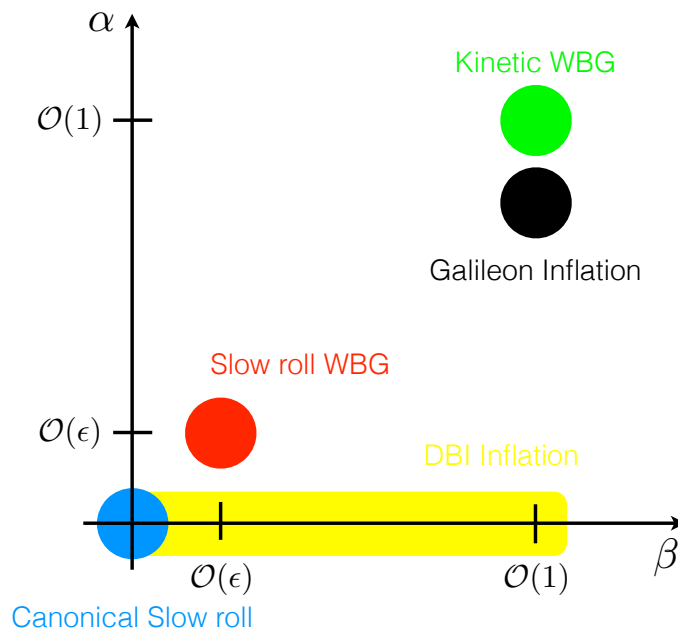


Figure 1: Various single-field models of inflation in the α - β plane.

The cartoon in Fig. 1 illustrates the regions in the α - β plane, occupied by the theories that we consider in what follows.

With the above remarks in mind, it is of some interest to study the observational status of the general parameter space defined by³

$$0 \lesssim \{ |\alpha|, |\beta|, |\gamma|, |\delta| \} \lesssim \mathcal{O}(1) . \quad (5)$$

This parameter space, as we argue in the next section, captures a great majority of existing single-field inflationary models and one purpose of the present paper is precisely to explore it in the light of the present experimental results.

As a technical, but crucial remark, we note that for the most general values of the parameters from the range (5), the so-called *decoupling limit* (DL) analysis of perturbations is not available. Indeed, the DL focuses on the dynamics of the Goldstone boson of time translation symmetry breaking π , ignoring its mixings with the metric degrees of freedom (this takes advantage of what is a direct analog of the *Goldstone boson equivalence theorem* in massive spin-1 theories [11].) While in many cases this is an adequate approximation at the relevant energy scales (i.e. those of order the inflationary Hubble rate), we will see that close to saturating the upper bound in (5), mixings of π with gravity become order-one important – invalidating any DL calculation of the scalar perturbations’ properties. Since we will find that various theoretical/experimental constraints already cover much of the parameter space of interest, even order-one theoretical effects can be important for studying the viability of some of the theories we consider below.

³To be as general as possible in the analysis below we will abuse our theoretical expectations by allowing a slightly larger—order ten—upper bound on the dimensionless EFT parameters.

Our calculation of the full cubic scalar action in the EFT (1) comes with several interesting by-products. Namely, we find that for a certain part of the parameter space, not only is mixing with gravity non-negligible, but it turns out to completely dominate non-Gaussianity, giving rise to the relation

$$f_{\text{NL}} \propto \frac{1}{c_s^6} \quad (6)$$

in the limit of a small speed of sound c_s . Such an abrupt growth of f_{NL} would have never been seen from a decoupling limit perspective, which gives $f_{\text{NL}} \propto 1/c_s^2$ for the same region of the parameter space⁴ (the regime characterized by (6) is studied in great detail in Appendix C.) As interesting as it is, however, we unfortunately find that (6) can only be of academic interest: the parameter space giving rise to such values of f_{NL} is already ruled out by the existing experimental bounds on primordial gravitational waves.

Another interesting effect is the inverse-proportional growth (for non-zero γ and δ) of the amplitude of non-Gaussianity with a small tensor-to-scalar ratio r :

$$f_{\text{NL}} = \gamma \frac{c_s}{r} \frac{80}{81} \frac{\alpha^2 - 3\alpha + 2 - \varepsilon + 3(\alpha - 1)^2 c_s^2}{(\alpha - 1)^4} - \delta \frac{c_s^3}{r} \frac{80}{81} \frac{1}{(\alpha - 1)^3} + \dots, \quad (7)$$

where we have defined $\varepsilon \equiv -\dot{H}/H^2$. This formula, supplemented by the present limits on the primordial gravitational waves, will play an important role in constraining the theories under consideration.

As to the phenomenology, we find that the constraints on the effective theory (1) are rather robust: despite the apparent multitude of the EFT coefficients, basic theoretical considerations (such as the absence of instabilities and of superluminal scalar modes) and current limits on the primordial gravitational waves and on non-Gaussianity already limit most of the parameter space. These constraints operate in a coordinated way, ruling out complementary regions of the latter. Moreover, we find that the theoretical and experimental viability of a given region of the parameter space is to a great extent determined by the set of just three numbers characterizing the operators in (1), quadratic in perturbations: α , β and ε . At the quadratic level, (1) in fact captures all single-field models with scalar perturbations obeying the usual, phonon-like dispersion relations

$$\omega = c_s k, \quad (8)$$

at energy scales of order H , and our results apply to any theory with the latter property.

Current data still allows for an appreciable range of parameters for the EFTI (1), leaving room for detectable non-Gaussianity. Reducing the existing upper bound on the tensor-to-scalar ratio by less than an order of magnitude,

$$r < 10^{-2}, \quad (9)$$

⁴See Refs. [12, 13] for the DL treatment.

would however put the theories that predict the tensor and the scalar tilts of the same order ($|n_T| \sim |n_s - 1| \sim \epsilon$) in a serious tension with experiment. The model that has a slightly better chance of surviving the bound (9) is slow-roll inflation with weakly broken galileon symmetry (SRWBG) of Ref. [14]. Unlike the canonical slow-roll models with plateau-like potentials famously consistent with (9), this model can be consistent with the latter constraint even if driven by a convex potential, as well as give rise to somewhat strongly coupled and highly non-Gaussian ($|f_{\text{NL}}| \sim 1 - 20$) scalar perturbations [14].

Set aside inflationary model-building, the results of this paper can be considered an extension of the calculations by Maldacena [15] and Chen et al. [16] of the full cubic action—including mixing with gravity—for the comoving curvature perturbation ζ . The first of these references has dealt with the case of canonical slow-roll inflation; Ref. [16], on the other hand, has generalized the analysis to the case when arbitrary terms of the form $P(\phi, (\partial\phi)^2)$ are present in the inflaton action. This is equivalent to considering operators of the form $(\delta N)^n$ in the EFT of inflationary perturbations. Here we further generalize the analysis of Chen et al. to the case when the operators of the form $(\delta N)^n \delta E$ become relevant in the unitary gauge EFT, which is naturally true for the broad class of theories with weakly broken galileon invariance. A calculation somewhat related to ours (although performed in a different language) has been carried out in Ref. [17]. That reference has however concentrated on the concrete model of *G-inflation*, and the corresponding calculation is less general than the effective field theory treatment we adopt here. We find the EFT approach rather convenient since it captures an overwhelming majority of the existing single-field models, as well as it makes transparent the above-described interesting features in Eqs. 6 and (7). Our presentation is close in spirit to those of [15, 16], and in the appropriate limits our results are in agreement with the results of these references, as well as of Ref. [17].

The rest of the paper is organized as follows. In Section 2 we study the dynamics of scalar perturbations in the theory specified by the action (1), and identify the full set of regions in the parameter space characterized by significantly subluminal/non-Gaussian scalar perturbations. Furthermore, we categorize the list of existing, as well as novel, models of inflation with large non-Gaussianity according to which one of these regions they fall into. In Section 3, we explore various constraints—both theoretical and experimental—that these theories are subject to. Finally, in Section 4, we conclude. Various technical details, that would overwhelm the main body of the text, are collected in the three appendices.

2 Dynamics of scalar perturbations

We start out with a brief account of the full analysis—including mixings with gravity—of scalar perturbations in the theory specified by the action (1). To this end, we closely follow and generalize Maldacena’s calculation [15] for canonical slow-roll inflation (see also Ref. [16].)

The gauge freedom that remains after fixing to the unitary gauge, $\delta\phi(x, t) = 0$, can be used [15] so as to put the three-dimensional metric into the following form, $g_{ij} = a^2 e^{2\zeta} (\delta_{ij} + h_{ij})$.

In this gauge, the scalar and tensor perturbations are captured by ζ and h_{ij} respectively (the lapse and shift variables are non-dynamical, and can be expressed in terms of the rest of the degrees of freedom using their equations of motion.) We will be exclusively interested in the n -point functions of ζ , so we ignore h_{ij} altogether in the remainder of this work. Integrating out the (scalar parts of) perturbations δN and $N_i \equiv \partial_i \psi$ from the Hamiltonian and momentum constraint equations, one obtains the effective action for the only remaining scalar degree of freedom, ζ . The procedure is quite tedious, but straightforward, and is outlined in Appendix A; the final result for the quadratic action reads

$$S^{(2)} = \int d^4x a^3 \mathcal{N} \left[\dot{\zeta}^2 - c_s^2 \frac{(\partial_i \zeta)^2}{a^2} \right], \quad (10)$$

where the explicit expressions for the kinetic normalization factor \mathcal{N} and the speed of sound c_s are given in Eqs. (35) and (36) of Appendix A. In terms of the dimensionless parameters of Eq. (5), these are

$$\begin{aligned} \mathcal{N} &= M_{\text{Pl}}^2 \frac{3\alpha(\alpha - 2) + \beta + \varepsilon}{(\alpha - 1)^2}, \\ c_s^2 &= \frac{\alpha(1 - \alpha) + \varepsilon}{3\alpha^2 - 6\alpha + \beta + \varepsilon}, \end{aligned} \quad (11)$$

We have neglected the $\partial_t \hat{M}_1^3$ term in the expression for c_s^2 , Eq. (36). More generally, we will neglect time derivatives of all the free EFT coefficients (in the second line) of the action (1) throughout this work⁵. This assumption will not change our conclusions in any appreciable way, and we choose to make it in order to keep the presentation as simple as possible.

In what follows we will be mostly exploring the properties of the theory in the $\alpha - \beta$ plane, fixing the rest of the parameters to some constant values. The reason is that it is precisely these two parameters—along with the slow-roll parameter ε —that determine the properties of the quadratic perturbation Lagrangian, see Eqs. (34) and (11). As a result, two out of the three constraints that we'll impose below, namely those stemming from stability/subluminality and the tensor-to-scalar ratio, are unambiguously determined by α , β and ε . As to the third constraint arising due to non-Gaussianity, the latter three parameters do also contribute to f_{NL} . Whenever this contribution becomes too large, one can turn on non-zero values for γ and δ to bring f_{NL} back within the observational limits. However, barring such an adjustment of parameters, it is a rather interesting fact that the set of just three numbers, α , β and ε , can tell us a great deal about the theoretical and experimental status of an overwhelming majority of single-field inflationary theories.

In fact, the latter set of parameters determines the phenomenology of any single-field model of inflation in which the scalar perturbations are characterized by the usual, phonon-like, dispersion relation (8) at freezeout of the CMB modes. The reason is as follows. At the quadratic

⁵One can expect that the time dependence of any EFT coefficient M^n in (1) satisfies $\partial_t M^n \sim \varepsilon' H M^n \ll H M^n$ on a quasi-de Sitter space, where ε' is slow-roll suppressed. As a result, the effects associated with non-zero time derivatives of these coefficients are generically suppressed with respect to their leading-order effects. We have explicitly checked this fact for the cases we consider below.

order in perturbations, there are only two additional operators one can add to the Lagrangian (1) to be consistent with Eq. (8). These are $\sqrt{-g} (\delta E_j^i \delta E_i^j - (\delta E)^2)$ and $\sqrt{-g} \delta N {}^{(3)}R$. Both of these operators, however, are redundant and can be removed by a perturbative field redefinition [18], so at least the quadratic piece of our action (1) is very generic. Since, as remarked above, the majority of the constraints on the EFT (1) stem precisely from the *quadratic* Lagrangian, our analysis should (at least qualitatively) capture phenomenology of any model satisfying (8).

Particularly interesting regions in the parameter space are the ones that correspond to the (squared) speed of sound, c_s^2 , becoming small. The reason is that this generically enhances the couplings of scalar perturbations, giving rise to fairly large values of non-Gaussianity, potentially relevant for observations. One can see, from Eq. (11), that there are a number of ways to make the sound speed suppressed. We will divide these into the following broad classes, which encompass most of the single-field models capable of generating large non-Gaussianity.

2.1 DBI, and related models

In the unitary gauge, these models give rise to the following relations between the dimensionless parameters [6, 16]

$$\{\alpha, \gamma\} \sim 0, \quad \{|\beta|, |\delta|\} \gtrsim \varepsilon, \quad (12)$$

implying the following behaviour of the amplitude of non-Gaussianity for strongly subluminal perturbations [16]

$$f_{\text{NL}} \propto \frac{1}{c_s^2}. \quad (13)$$

The two non-zero coefficients β and δ in fact parametrically exceed the slow-roll parameter ε in the small c_s^2 limit: $\beta \propto \varepsilon/c_s^2$, and $\delta \propto \varepsilon/c_s^4$. However, current experimental bounds imply $c_s^2 \gtrsim \varepsilon$ for the DBI model [19], so perhaps the optimal values for these parameters to keep in mind are

$$|\beta| \lesssim 1, \quad |\delta| \sim \frac{\beta}{c_s^2}. \quad (14)$$

2.2 G-inflation/Galileon inflation

In this category, we collect inflationary theories characterized by the dimensionless parameters of (1) satisfying the following conditions

$$\{|\beta|, |\gamma|, |\delta|\} \sim 1, \quad |\alpha| < 1, \quad (15)$$

so that the subluminal limit corresponds to

$$c_s^2 \sim \frac{\alpha}{\beta} < 1. \quad (16)$$

This can be the case in G-inflation [20], Galileon inflation [13] and, more generally, in the kinetically driven phase of theories with weakly broken galileon symmetry [9]. For a suppressed speed of sound, the amplitude of non-Gaussianity grows similarly to DBI models

$$f_{\text{NL}} \propto \frac{1}{c_s^2} . \quad (17)$$

It has been noticed by Burrage et al. [13], however, that in the unitary gauge, the most general theory of Galileon inflation introduces extra cubic operators on top of the ones present in (1), such as $\sqrt{-g} \delta N (\delta E_j^i \delta E_i^j - (\delta E)^2)$ for example (with an order-one coefficient in Planck units), and these can result in a faster growth

$$f_{\text{NL}} \propto \frac{1}{c_s^4} . \quad (18)$$

We have not included such operators in our analysis. The major reason is that, as we will see shortly, constraints on various inflationary models described by the EFT (1) are already quite strong regardless of the cubic operators; including the latter can loosen the constraints due to non-Gaussianity⁶, but only at an expense of tuning/cancellations. As to the rest of the constraints of the next section, they do apply equally well to the most general theory of G-/Galileon inflation, as discussed above Eq. (8).

2.3 Kinetically driven inflation with WBG symmetry

This model is characterized by

$$\alpha \simeq 1 , \quad \{ |\beta|, |\gamma|, |\delta| \} \sim 1 , \quad (19)$$

and arises in the context of general theories with weakly broken galileon symmetry [9], of which the *Covariant Galileon* [21] behind Galileon inflation is a particular case. The values of the parameters in (19) can in fact also arise in Galileon inflation [13]; unfortunately, however, the corresponding regime can not be captured by the decoupling limit analysis adopted in the latter reference. We have chosen to present this theory as a separate class, since it gives rise to the fastest growth of non-Gaussianity with the small speed of sound,

$$f_{\text{NL}} \propto \frac{1}{c_s^6} , \quad (20)$$

well within the regime of validity of the low energy EFT. We give a detailed account of the kinetically driven inflation with weakly broken galileon symmetry (KWBG) in Appendix C. The expression in Eq. (20) arises as a result of subtle effects associated with mixing of the adiabatic perturbations with the metric degrees of freedom, partially explaining why it has gone unnoticed in the literature.

⁶Being manifestly at least cubic in perturbations, these operators only affect the bispectrum from the set of physical quantities we consider below.

2.4 Slow-roll inflation with WBG symmetry

This class of theories was introduced in Ref. [14] and gives rise to the values of the dimensionless parameters of the EFT (1) suppressed by slow-roll

$$\{ |\alpha|, |\beta|, |\gamma|, |\delta| \} \sim \varepsilon. \quad (21)$$

The speed of sound of scalar perturbations is generically order-one (but not necessarily strictly one, in contrast to canonical slow-roll inflation.) The background evolution, along with all background characteristics (spectral tilt, number of e-folds from freezeout of the CMB modes until the end of inflation, spectrum of gravitational waves, etc.) are all parametrically similar to garden-variety slow-roll models. What's different, though, is that the scalar perturbations can be more strongly coupled than in canonical slow-roll inflation – in a way that, nevertheless, allows to keep control over the derivative expansion. This leads to an amplitude of non-Gaussianity that grows like

$$f_{\text{NL}} \propto \frac{1}{c_s^4} \quad (22)$$

in the subluminal limit. Note that, while this behaviour is similar to (18), the underlying models are very different: the SRWBG model is a minimal deformation of slow-roll inflation, unlike Galileon inflation which describes a *kinetically-driven* background. Moreover, as already mentioned above, the version of Galileon inflation described by the EFT (1) in fact yields a DBI-like growth, Eq. (13), while the slow-roll theories with WBG invariance lead to (22) already within the realm of the EFT (1) – i.e. without the need of having extra cubic operators. The strong dependence of non-Gaussianity on the speed of sound allows to generate appreciable values for f_{NL} even for mildly subluminal perturbations, arising for

$$\alpha \simeq -\varepsilon \quad (23)$$

in the SRWBG model (see Eq. (11).)

Having classified various alternatives to canonical slow-roll inflation according to the way they generate large non-Gaussianity, we proceed to explore the theoretical and experimental constraints on the general EFT parameter space (5) in the next section.

3 Constraints

There is a number of constraints—both theoretical and experimental—that the models 2.1 through 2.4 discussed in the previous section are subject to. Above all, there is a constraint expressing the absence of negative norm states (or, alternatively, boundedness from below of the Hamiltonian) and of gradient instability. Theories, that do not satisfy this constraint can hardly be made sense of. A somewhat less sharp constraint comes from demanding the absence of superluminal scalar perturbations. While it is not fully clear whether superluminality within a low-energy EFT is unconditionally unacceptable, there are good reasons to believe that

at the very least it is inconsistent with the standard properties (Lorentz-invariance, locality, analyticity, etc.) of a hypothetical UV completion [22]. To be on the safe side, we will thus demand that the scalar excitations are subluminal as well. The above considerations then summarize into the following conditions on the dimensionless parameters of the theory

$$\mathcal{N} > 0, \quad 0 < c_s^2 \leq 1. \quad (24)$$

Furthermore, an important role for our analysis will be played by the current limits on the amplitude of primordial gravitational waves. The scalar-to-tensor ratio, r , can be readily read off the quadratic ζ action, Eq. (34),

$$r = 16 \frac{\mathcal{N} c_s^3}{M_{\text{Pl}}^2} = 16 \frac{\varepsilon + \alpha - \alpha^2}{(\alpha - 1)^2} \sqrt{\frac{\varepsilon + \alpha - \alpha^2}{3\alpha^2 - 6\alpha + \beta + \varepsilon}}. \quad (25)$$

In the slow-roll limit, $\alpha = \beta = 0$, this reduces to the familiar expression $r_{\text{sr}} = 16\varepsilon$, while for DBI inflation ($\alpha = 0$), Eqs. (25) and (11) yield, $r_{\text{DBI}} = 16\varepsilon c_s$.

Last but not least, we will impose the experimental limits on primordial non-Gaussianity. The full calculation of the scalar bispectrum for the theory (1) is presented in Appendices (A) and (B), and we will not reproduce it here. We define the (shape-independent) measure for the strength of non-Gaussian effects, f_{NL} , in the following way

$$f_{\text{NL}} = \frac{5}{18} \frac{B_\zeta(k, k, k)}{P_\zeta(k)^2}, \quad (26)$$

where the bispectrum B_ζ can be found from the three-point function of ζ

$$\langle \zeta(\vec{k}_1) \zeta(\vec{k}_2) \zeta(\vec{k}_3) \rangle = (2\pi)^3 \delta^{(3)} \left(\sum_i \vec{k}_i \right) B_\zeta(k_1, k_2, k_3). \quad (27)$$

As to the shape of non-Gaussianity, it is generically well-approximated by the equilateral one [23] in the theory (1).

The precise expression for f_{NL} in terms of the dimensionless parameters of the action (1) (including the slow-roll parameter ε) is not particularly illuminating. As discussed in the previous section, in different regions of the parameter space corresponding to significantly subluminal scalar perturbations, f_{NL} acquires simple leading behaviour of the type $f_{\text{NL}} \propto 1/c_s^{2p}$, with $p = 1, 2$ or 3 . Moreover, for non-zero γ and δ , there is a " $f_{\text{NL}} \propto 1/r$ effect", mentioned below Eq. (7).

We will ultimately project the parameter space onto the $\alpha - \beta$ plane, so some input regarding the magnitude of the slow-roll parameter ε is needed. The measured tilt of the primordial scalar spectrum n_s suggests that $\varepsilon \lesssim |n_s - 1| \sim 10^{-2}$. The latter bound is saturated for many inflationary models – e.g. those with convex potentials. On the other hand, there are models characterized by plateau-like potentials such as Starobinsky's R^2 inflation [24], or the so-called IR DBI inflation [25], where ε can be much smaller than the scalar tilt. In order to capture

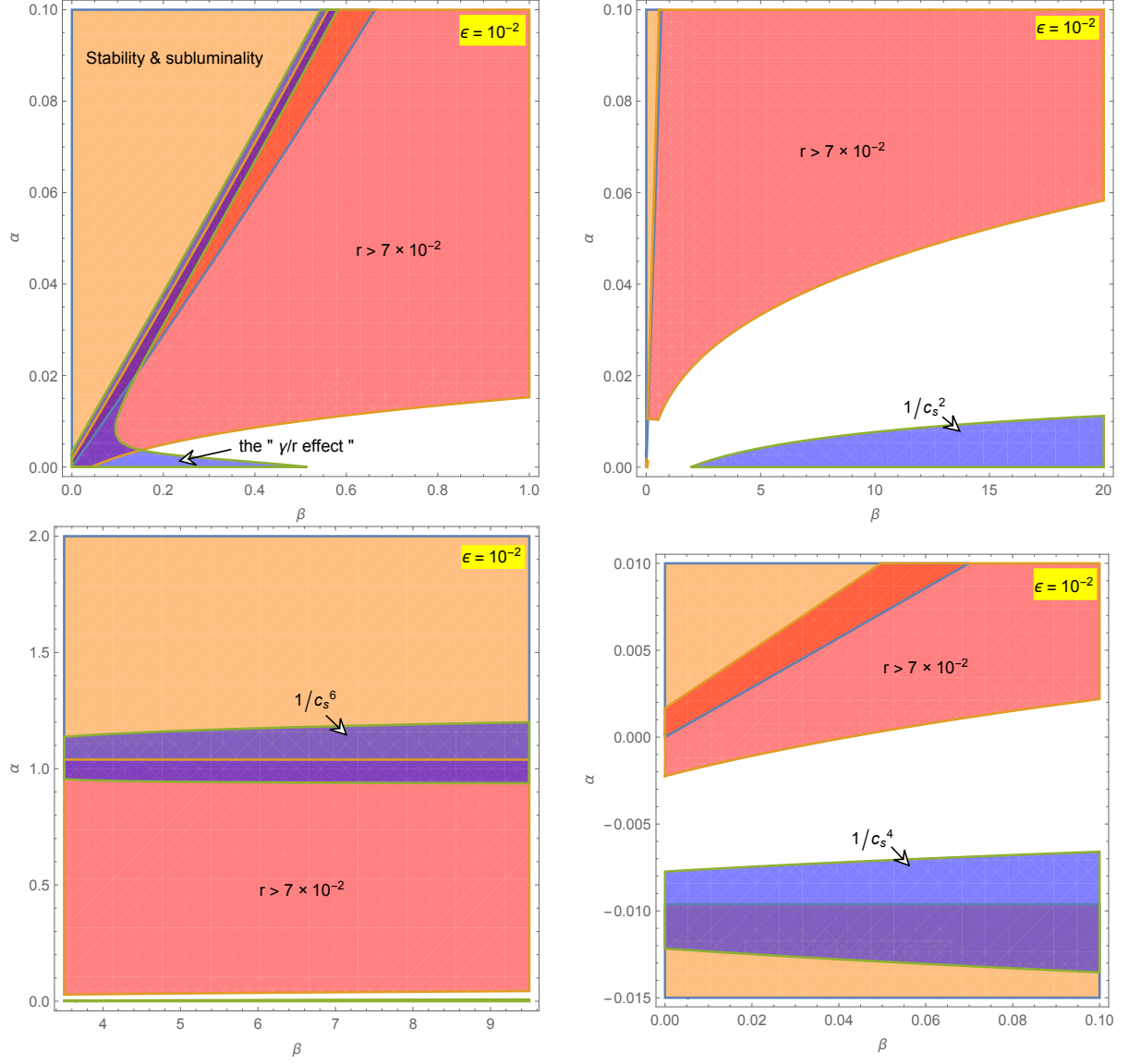


Figure 2: Various regions of the parameter space, excluded by the requirements of stability & subluminality (orange), current limits on the amplitude of primordial gravitational waves (red), and on non-Gaussianity (blue). The slow-roll parameter has been fixed to $\epsilon = 10^{-2}$, and the red and blue bands correspond to regions excluded respectively by the bounds $r < 0.07$ and $-50 < f_{\text{NL}} < 50$. The parameters γ and δ have been chosen to vanish everywhere except the upper left panel, where they have been set to $\gamma = \delta = 5$.

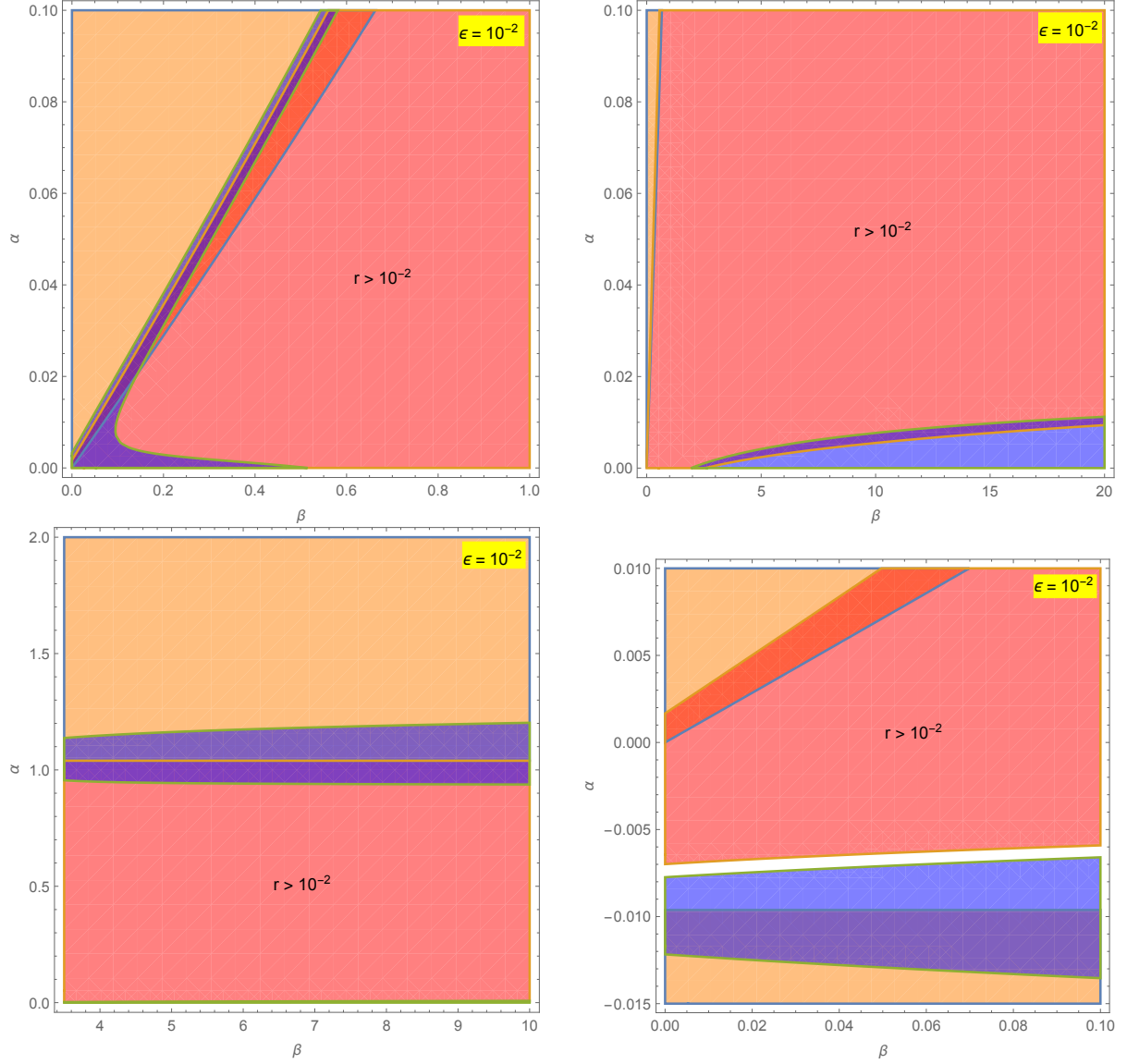


Figure 3: Here we illustrate the exact same situation as in Fig. 2, with the only difference in the exclusion from the tensor-to-scalar ratio: the red band now corresponds to the regions, where $r > 10^{-2}$.

both classes of models, we will assume two values for the slow roll parameter in our analysis: $\epsilon = 10^{-2}$ and $\epsilon \sim 0$ (the latter precisely defined below.)

The three constraints discussed above lead to an interesting interplay, in many cases excluding complementary regions of the parameter space. Consider, for example, a DBI-like model with a generic power-law potential, so that $\epsilon \sim 10^{-2}$. The tensor-to-scalar ratio, $r_{\text{DBI}} = 16\epsilon c_s$, has to be below ~ 0.1 according to the current experimental limits [19], requiring a somewhat suppressed speed of sound. On the other hand, significantly suppressing c_s , one runs into tension with the current limits on non-Gaussianity, in accord with Eq. (13). The precise constraint for DBI models is $c_s \gtrsim 0.1$ [26]. This means that measuring $r \sim 10^{-2}$ would rule out the given class of theories. In contrast, DBI theories driven by plateau-like potentials like the IR model of Ref. [25], are characterized by $\epsilon \ll 10^{-2}$ and therefore have a better chance of being consistent with a small tensor-to-scalar ratio.

The examples of exclusion plots for the general parameter space of interest are shown on Figs. 2 through 5. On the first two of these figures, we assume $\epsilon = 10^{-2}$, while the last two correspond to $\epsilon \sim 0$. Moreover, we require f_{NL} to be in the range $-50 < f_{\text{NL}} < 50$, motivated by the current limits on equilateral non-Gaussianity [26]. The orange, red and blue regions depict parts of the parameter space excluded by instabilities and/or superluminality, limits on the tensor-to-scalar ratio, and on non-Gaussianity respectively.

3.1 Models with $\epsilon \sim |n_s - 1|$

On Fig. 2 the red regions are the exclusion bands due to the present 95% C.L. bound on the amplitude of the primordial gravitational waves, $r < 0.07$ [27]. One can see, that the data prefers significantly suppressed α , effectively ruling out inflationary theories with $\alpha \gg 10^{-2}$ (this is a rather general result, true for all cases that we consider below ⁷.)

For α much smaller than ϵ , the boundary between the DBI and G-/Galileon inflation is blurred. An important discriminant that remains, though, is the fact that γ can be much larger in the latter class of models. For this reason, we have chosen $\gamma = \delta = 5$ in the upper left panel, which results in an additional exclusion region in G-/Galileon inflation due to large non-Gaussianity stemming from the "1/r" effect of Eq. (7) (from the two terms in this equation, only the one proportional to γ contributes significantly.) Had we chosen $|\gamma| \ll 1$ as in DBI inflation, this band would have completely disappeared from the plot. The regions of the α - β plane explored in the rest of the panels on Fig. 2 are not affected by γ and δ for reasonable values of these parameters⁸ so we have set them to zero everywhere except the upper left one.

The upper right panel of Fig. 2 shows the blue exclusion region due to non-Gaussianity in the $f_{\text{NL}} \propto 1/c_s^2$ regime, characteristic of the models 2.1 and (2.2). The blue band shown here

⁷The allowed region can in fact reach out to $\alpha \simeq 0.1$, but this only happens for $\beta \sim 50$, casting shadow on the quantum stability of the corresponding theories.

⁸For the upper right and the lower left panels, even setting $\gamma \sim \delta \sim 10$ has little effect on the exclusion regions. The lower right panel corresponds to the model 2.4 (SRWGB), where both of these parameters are naturally of order ϵ and lead to negligible effects.

appears for larger values of β , where the speed of sound becomes small enough (see Eq. (11))

$$c_s^2 \sim \frac{\mathcal{O}(\alpha, \varepsilon)}{\beta}, \quad (28)$$

so as to trigger the growth of f_{NL} according to Eqs. (13) and (17).

Another part of the parameter space, corresponding to the model 2.3 (KWBG) is shown on the lower left panel. One can see, that the region excluded by non-Gaussianity due to the abrupt growth $f_{\text{NL}} \propto 1/c_s^6$ is concentrated around $\alpha \simeq 1$ (as expected from Eq. (19).) Unfortunately, this model is already ruled out by the limits on the primordial gravitational waves, combined with the theoretical requirements of stability and subluminality.

Finally, on the lower right panel, we zoom onto the parameter space corresponding to the slow-roll WBG model of Ref. [14], discussed in Sec. 2.4. One can see the exclusion band from non-Gaussianity around $\alpha \simeq -\epsilon$, corresponding to f_{NL} growing like $\sim 1/c_s^4$. Just like in DBI/G-/Galileon inflation, there remains an appreciable portion of the parameter space still allowed by our constraints with $r \lesssim 0.07$, including regions of the α - β plane characterized by detectable non-Gaussianity. We stress again, however, that the SRWBG model, being in a well-defined sense a minimal deformation of canonical slow-roll inflation, is very different from the rest of the models considered in Sec. 2.

While the current data still leaves some room for most of the models with $\varepsilon \sim |n_s - 1|$, the situation can change dramatically if the upper limit on r decreases to $r \lesssim 10^{-2}$ (which is less than an order of magnitude improvement in current precision.) The plots, corresponding to this case are shown on Fig. (3); one can see that the regions that were previously allowed are now fully covered by the exclusion bands from gravitational waves. The only region that still remains is a narrow band in the slow-roll model with weakly broken galileon invariance, shown on the lower right panel of Fig. 3.

The canonical models of slow-roll inflation sit at the origin of the α - β plane, and are of course not visible on our plots. Measuring $r \lesssim 10^{-2}$ would rule out most of these, with an exception of models with plateau-like potentials, such as Starobinsky's R^2 inflation [24]. In these models, the tilt of the scalar spectrum is mostly determined by the second slow-roll parameter, $\eta_V \equiv M_{\text{Pl}}^2 V''/V$, so that ε can be much smaller than $|n_s - 1|$ to suppress the tensor-to-scalar ratio. Needless to say, falling into the category of canonical slow-roll theories, R^2 inflation predicts undetectable non-Gaussianity, $f_{\text{NL}} \sim 10^{-2}$ [15]. In contrast, (the non-canonical) slow roll inflation with WBG symmetry, *even if driven by the simplest convex potentials* (with $\varepsilon \sim |n_s - 1|$), does possess a parameter space consistent with tensor-to-scalar ratios as small as $r \lesssim 10^{-2}$, as seen from the lower right panel of Fig. 3. Moreover, close to the blue non-Gaussianity exclusion band, this model can generate detectable (equilateral) non-Gaussianity, $|f_{\text{NL}}| \lesssim 50$ [14].

3.2 Models with $\varepsilon \ll |n_s - 1|$

It is well-known that the tensor-to-scalar ratio can be significantly suppressed (and therefore the bounds coming from this observable ameliorated) in models where the variation of the

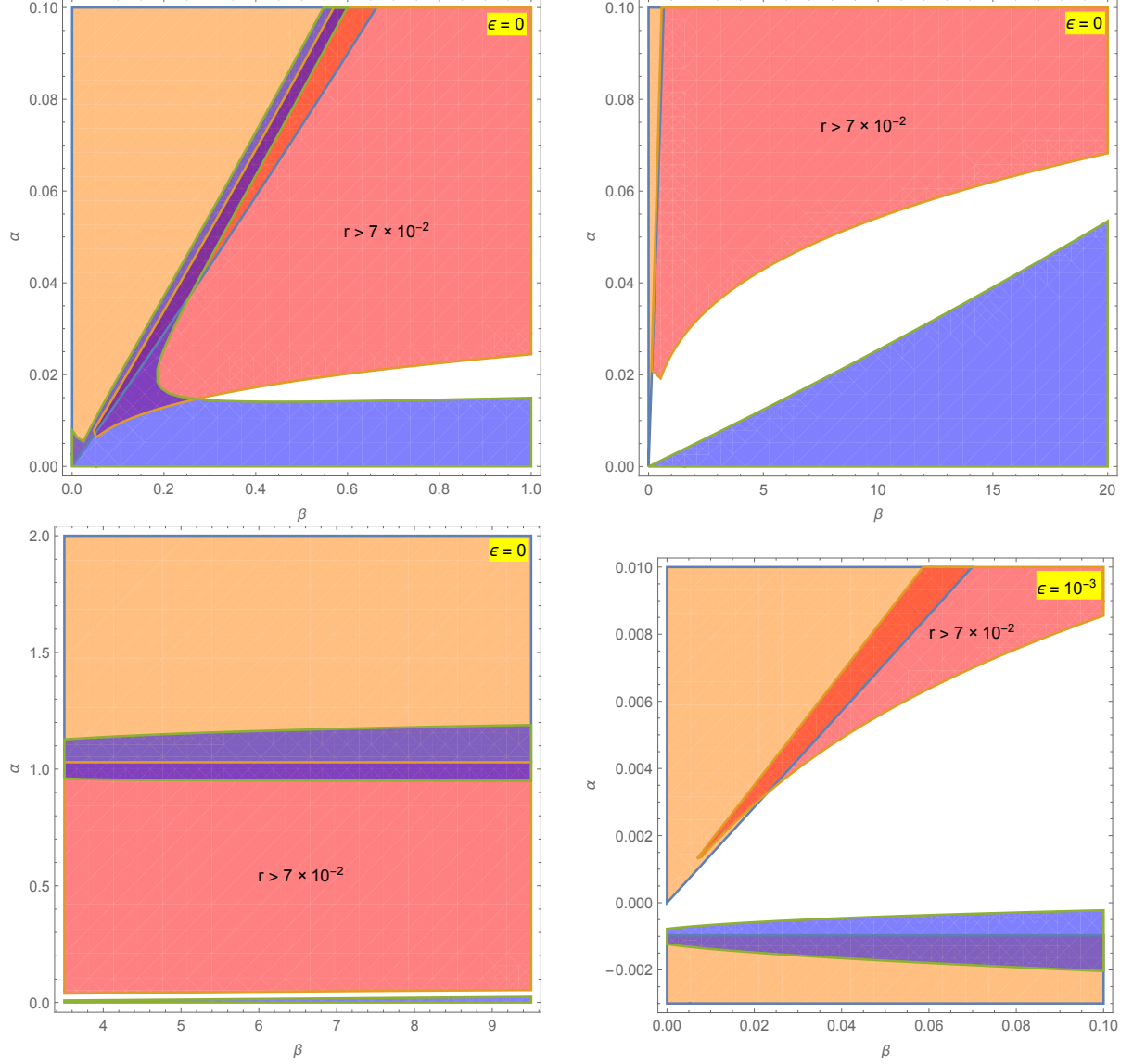


Figure 4: Various regions of the parameter space corresponding to models discussed in Sec. (2). The red band shows the regions excluded by requiring $r < 7 \times 10^{-2}$, while $\epsilon = 0$ in all panels but the lower right one, which has $\epsilon = 10^{-3}$.

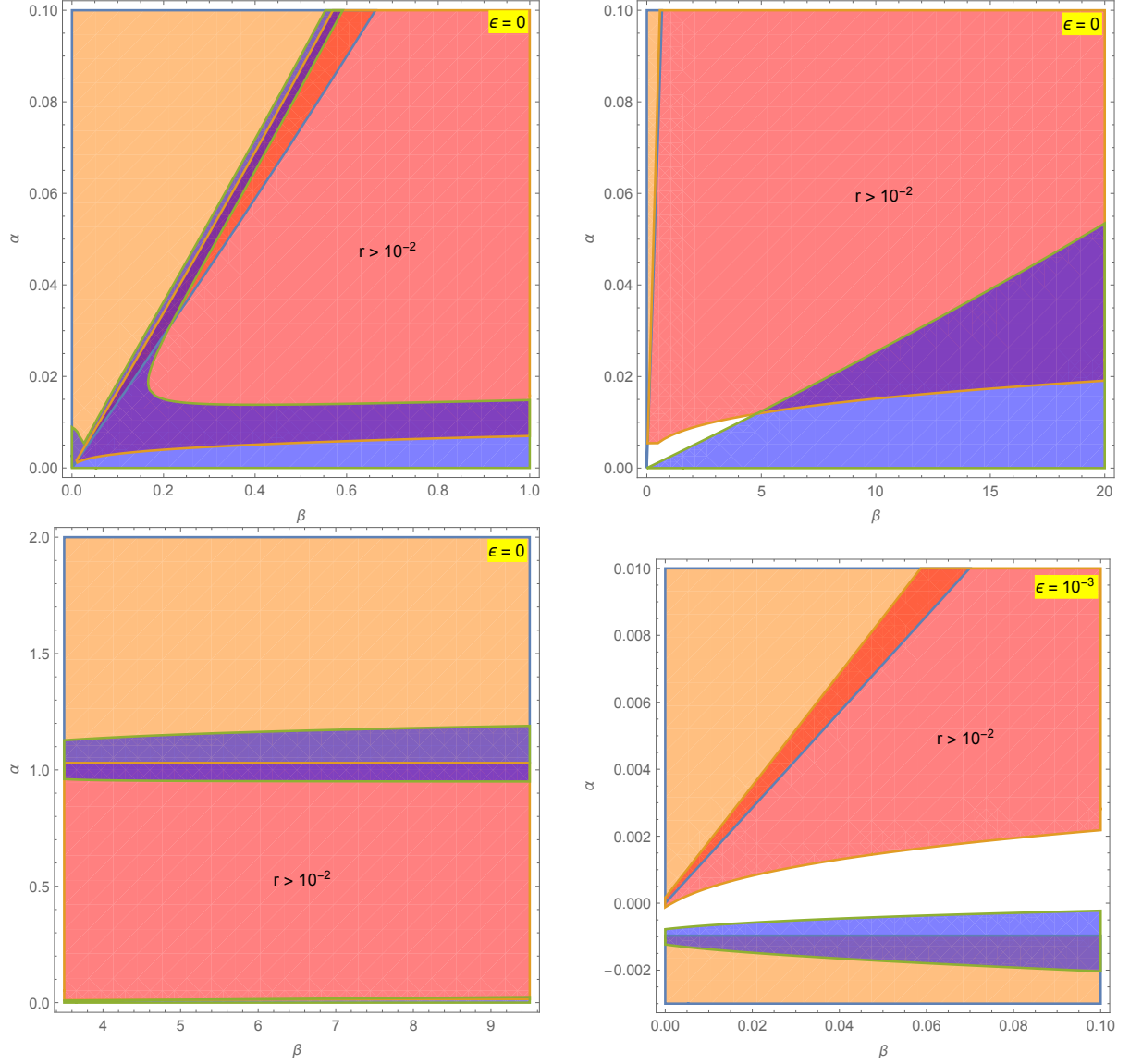


Figure 5: Here we illustrate the exact same situation as in Fig. 4, with the only difference in the exclusion from the tensor-to-scalar ratio: the red band now corresponds to the regions, where $r > 10^{-2}$.

inflationary Hubble rate does not significantly contribute to the scalar tilt. A famous example are theories (slow-roll or not) driven by plateau-like potentials. To capture this class of models, we repeat the analysis of the previous subsection setting now $\varepsilon = 0$ for all plots⁹ but that corresponding to slow-roll WBG inflation (in which the regime with large non-Gaussianity crucially depends on the presence of a non-zero slow-roll parameter ε , see Eq. (23).) In the latter case we set $\varepsilon = 10^{-3}$. Apart from these modifications, the figures 4 and 5 correspond to the exact same choices of parameters as in Figs. 2 and 3 respectively.

The situation for $\epsilon \ll |n_s - 1|$ is qualitatively similar to the previous case ($\epsilon \sim |n_s - 1|$.) The current data still allows parameter space, consistent with the existing bounds on r and corresponding to measurable non-Gaussianity. However, improving the limits on the amplitude of primordial gravitational waves could still induce dramatic changes. One novelty compared to the case of the previous subsection is that for $\epsilon \sim 0$, there would still remain a small allowed parameter space for DBI/G-/Galileon inflation even if $r < 10^{-2}$ (see the upper right panel of Fig. 5). Moreover, a much larger fraction of the allowed parameter space would survive the $r < 10^{-2}$ bound in the slow roll WBG model, as seen from the lower right panels of Figs. 4 and 5.

4 Discussion and future directions

In this paper, we have explored a number of constraints—both theoretical and experimental—on alternative theories to canonical slow-roll inflation, described by the general unitary-gauge action (1). The latter action captures the great majority of single-field models motivated by considerations of theoretical consistency (in particular, stability under quantum corrections) and phenomenological interest (possibility to generate non-Gaussianity, detectable by current or near-future experiments.) The list of models covered by our analysis (1) includes, but is not limited to, DBI inflation [5, 6], G-inflation [20], Galileon inflation [13], the recently-proposed slow-roll inflation with weakly broken galileon symmetry [14], etc.

We have found that the requirement of the absence of instabilities and/or superluminal propagation, the current experimental limits on the tensor-to-scalar ratio and on non-Gaussianity result in an interesting interplay, ruling out complementary parts of the parameter space (see Figs. 2 through 5.) The existing upper bound on the amplitude of primordial gravitational waves, $r \lesssim 0.07$, still allows for an appreciable range of parameters for most of the above models, leaving room for detectable non-Gaussianity. The only theory which is already ruled out is the KWBG model 2.3, characterized by an abrupt growth of non-Gaussianity, $f_{\text{NL}} \propto 1/c_s^6$, in the subluminal limit (note, that this model is excluded in a quite non-trivial way, however – see the lower left panel of Fig 2.)

⁹Ideally speaking, DBI inflation corresponds to a vanishing coefficient α . Setting $\varepsilon = 0$ then sends the speed of sound to zero, or equivalently, f_{NL} to infinity, ruling out the DBI models with vanishing ε . This can be seen e.g. on the upper right panel of Fig. 4, where the non-Gaussianity exclusion band covers the whole $\alpha = 0$ axis. Of course, in a more realistic situation with a small but non-zero ε , an allowed region with $\alpha = 0$ and a small-enough β opens up.

Reducing the upper bound on the tensor-to-scalar ratio down to $r < 10^{-2}$ would put all of the above-considered theories in some tension with experiment. This situation is depicted on Figs. 3 and 5. The former figure corresponds to models with $\varepsilon \sim |n_s - 1|$, in which case the only allowed region that has a chance of surviving our constraints (shown on the lower right panel) belongs to slow-roll inflation with weakly broken galileon symmetry of Ref. [14]. Fig. 5, on the other hand, shows that the theories characterized by $\varepsilon \ll |n_s - 1|$ have more chance of being consistent with small values of the tensor-to-scalar ratio. The SRWBG model seems to work slightly better compared to the other models also in this case, retaining a bigger fraction of its allowed parameter space with the decrease of the upper bound on r . We thus find that, in addition to models driven by plateau-like potentials famously consistent with (9), there would still remain a chance for an alternative single-field model driven by a *convex potential* ($\varepsilon \sim |n_s - 1|$), even if r is measured to be as small as in Eq. (9). While the latter model is slow-roll at the level of the background evolution, there are quantitative differences from the standard models with a canonical kinetic term. Most dramatically, the dynamics of perturbations is rather different in this model, allowing for somewhat strongly coupled and highly non-Gaussian scalar perturbations ($f_{\text{NL}} \sim 1 - 20$) [14].

Our analysis is rather general but it is not without loopholes, of course. We have argued above that, at the quadratic order, (1) is the most general action that captures theories characterized by scalar perturbations with usual, phonon-like dispersion relations (8). However, there exist models such as *ghost inflation* [28, 29, 30], where $\alpha \sim 0$ and the background describes a perfect de Sitter space (i.e. $\varepsilon = 0$), so that the scalar speed of sound vanishes at the zeroth order (consistent with our general expression for c_s^2 , Eq. (11).) In such a case, one ought to consider effects of higher-derivative operators in (1)—e.g. δK^2 —that will dominate the gradient energy of the scalar modes at horizon crossing (i.e., at characteristic frequencies of order $\omega \sim H$.) This results in a rather different, $\omega = k^2/M$ dispersion relation with M some cutoff scale, which would presumably modify our analysis. It would be interesting to see how our conclusions are affected in this class of models.

While our theoretical calculation is complete, a precise analysis of the data on non-Gaussianity has remained outside of the scope of this work. The reason is that while the shape of non-Gaussianity generated by the effective theory (1) is always close to equilateral, it of course does not generically coincide with the templates constrained by experiments; therefore, extracting precision experimental limits is not straightforward. One possible improvement of our results (which we believe won't bring significant changes to our conclusions) would precisely be to do a precision analysis of the existing data on the scalar bispectrum in the context of our approach.

Most importantly, however, our analysis reveals an interesting interplay between various theoretical and observational constraints that exclude complementary parts of the parameter space in general inflationary effective theories like the one of Eq. (1). We believe that further developing this approach via incorporating more general effective theories and/or more precise handling of the data could be a useful step in better understanding the ultimate physics behind inflation.

Acknowledgements

It is a pleasure to thank Paolo Creminelli, Gregory Gabadadze, Riccardo Rattazzi and Sergey Sibiryakov for valuable discussions and comments. The work of E.T. is supported in part by MIUR-FIRB grant RBFR12H1MW.

A The cubic ζ action

In this appendix, we derive the cubic Lagrangian for the comoving curvature perturbation ζ for a general effective theory described by the action (1). The presentation closely follows and generalizes Maldacena's calculation of an analogous Lagrangian for canonical slow-roll inflation [15], see also Ref. [16] for a generalization. We will work with the ADM variables,

$$ds^2 = -N^2 dt^2 + g_{ij}(N^i dt + dx^i)(N^j dt + dx^j)$$

and denote by ${}^{(3)}R$ the three-dimensional curvature corresponding to the metric g_{ij} induced on equal-time hypersurfaces, while the tensor E_{ij} is related to their extrinsic curvature K_{ij} as follows

$$E_{ij} = NK_{ij} = \frac{1}{2}(\partial_t g_{ij} - \nabla_i N_j - \nabla_j N_i) . \quad (29)$$

For a given inflationary spacetime, the first line of (1) is completely fixed by the background dynamics (i.e. by the Hubble rate and its first time derivative), while the second line (which starts at least quadratic in metric perturbations) is *a priori* unconstrained. For a concrete inflationary model, the coefficients M_2^4 , M_3^4 , etc. will depend on the parameters of the underlying theory. A unitary-gauge action of the form (1) propagates a single scalar mode in addition to the usual transverse-traceless graviton. We exclusively concentrate on the dynamics of the former mode in this appendix, and ignore the tensor and the (non-dynamical) vector degrees of freedom from now on. The scalar mode is conveniently studied in the gauge defined by $g_{ij} = a^2(t)e^{2\zeta}\delta_{ij}$, where ζ has the meaning of the curvature perturbation of spatial hypersurfaces. In order to compute the cubic action for ζ in the theory (1), we proceed as follows. We first integrate out the perturbations of the lapse and shift variables, δN and $N_i \equiv \partial_i \psi$, from their respective equations of motion, and then plug the solutions back into (1) to find the action for ζ at the desired order. To determine the cubic action, we only need to solve for the lapse and shift to the *linear* order in ζ (see Ref. [16] for a discussion on this point.) The Hamiltonian and momentum constraint equations, obtained by varying the action w.r.t. N and N^j , are

$$\frac{M_P^2}{2} \left[{}^{(3)}R - \frac{1}{N^2} (E^{ij} E_{ij} - E^2) + \frac{2}{N^2} \dot{H} - 2(3H^2 + \dot{H}) \right] + M_2^4 \delta N - \hat{M}_1^3 \delta E = 0, \quad (30)$$

$$\hat{\nabla}_i \left[\frac{M_P^2}{N} (E_j^i - \delta_j^i E) - \hat{M}_1^3 \delta_j^i \delta N \right] = 0, \quad (31)$$

and one can readily find the linear solution

$$\delta N = \frac{2M_P^2}{2M_P^2 H - \hat{M}_1^3} \dot{\zeta} \quad (32)$$

$$\psi = -\frac{2M_P^2}{2M_P^2 H - \hat{M}_1^3} \zeta + \frac{2M_P^2 M_2^4 - 12M_P^2 H \hat{M}_1^3 + 3\hat{M}_1^6 - 4M_P^4 \dot{H}}{(2M_P^2 H - \hat{M}_1^3)^2} \left(\frac{\partial}{a}\right)^{-2} \dot{\zeta} \equiv C\zeta + \chi. \quad (33)$$

When plugged back into the original action (1), this yields the following quadratic Lagrangian for the comoving curvature perturbation

$$S^{(2)} = \int d^4x a^3 \mathcal{N} \left[\dot{\zeta}^2 - c_s^2 \frac{(\partial_i \zeta)^2}{a^2} \right], \quad (34)$$

where we have defined

$$\mathcal{N} \equiv M_P^2 \frac{-4M_P^4 \dot{H} + 2M_P^2 M_2^4 - 12M_P^2 H \hat{M}_1^3 + 3\hat{M}_1^6}{(2M_P^2 H - \hat{M}_1^3)^2}, \quad (35)$$

$$c_s^2 = \frac{4M_P^4 \dot{H} - 2M_P^2 H \hat{M}_1^3 + \hat{M}_1^6 - 2M_P^2 \partial_t \hat{M}_1^3}{4M_P^4 \dot{H} - 2M_P^2 M_2^4 + 12M_P^2 H \hat{M}_1^3 - 3\hat{M}_1^6}. \quad (36)$$

Obtaining the cubic-order action requires significantly more work. A straightforward expansion of (1) using the solutions (32) and (33) for the lapse and shift perturbations, yields the following result

$$\begin{aligned} S^{(3)} = \int d^4x \left\{ -a\mathcal{N}c_s^2 \zeta (\partial_i \zeta)^2 + a^3 \left[C\mathcal{N} + \frac{C^3}{M_{\text{Pl}}^4} \lambda \right] \dot{\zeta}^3 + 3a^3 \mathcal{N} \dot{\zeta} \dot{\zeta}^2 + \right. \\ \left. + \frac{M_P^2}{2a} (3\zeta + C\dot{\zeta}) [(\partial_i \partial_j \psi)^2 - (\partial^2 \psi)^2] - \frac{2M_P^2}{a} \partial_i \zeta \partial_i \psi \partial^2 \psi + \right. \\ \left. + aC^2 (2\hat{M}_1^3 - \hat{M}_2^3) \dot{\zeta}^2 \partial^2 \psi \right\}, \quad (37) \end{aligned}$$

where

$$\lambda \equiv -\frac{M_P^2}{2} \left[2M_P^2 (M_2^4 + M_3^4) - 3(2M_P^2 H - \hat{M}_1^3)(2\hat{M}_1^3 - \hat{M}_2^3) \right]. \quad (38)$$

It will prove convenient to recast the action (37) into a slightly different form. In doing so, we omit a number tedious but straightforward manipulations. Once the dust settles, one finds the following expression, equivalent to (37) up to a total derivative

$$\begin{aligned} S^{(3)} = \int d^4x \left\{ a^3 \left[\mathcal{N}C \left(1 + \frac{HC}{c_s^2} \right) - \lambda' \right] \dot{\zeta}^3 + \frac{a^3 \mathcal{N} (HC)^2}{c_s^2} \left(\varrho - 3 + \frac{3c_s^2}{(HC)^2} \right) \zeta \dot{\zeta}^2 \right. \\ \left. + a\mathcal{N} (HC)^2 \left(\varrho - 2s + 1 - \frac{c_s^2}{(HC)^2} \right) \zeta (\partial_i \zeta)^2 + 2a\mathcal{N} HC \dot{\zeta} \partial_i \zeta \partial_i \chi + \frac{a^3 \mathcal{N}}{2} \frac{d}{dt} \left(\frac{n(HC)^2}{c_s^2} \right) \zeta^2 \dot{\zeta} \right. \\ \left. + \frac{\mathcal{N}c_s^2}{2a} \partial_i \zeta \partial_i \chi \partial^2 \chi + \frac{\mathcal{N}c_s^2 + 2\hat{M}_1^3 C}{4a} \partial^2 \zeta (\partial_i \chi)^2 + \frac{\hat{M}_1^3 C^2}{a} \partial^2 \zeta \partial_i \zeta \partial_i \chi + \frac{\hat{M}_1^3 C^3}{2a} (\partial_i \zeta)^2 \partial^2 \zeta \right. \\ \left. + aC^3 (2\hat{M}_1^3 - \hat{M}_2^3) \dot{\zeta}^2 \partial^2 \zeta - g(\zeta) \frac{\delta L}{\delta \zeta} \right\}. \quad (39) \end{aligned}$$

In the last expression, we have defined a number of quantities

$$n = \frac{1}{H} \frac{d}{dt} \log(\mathcal{N}c_s^2), \quad \varrho = \frac{1}{H} \frac{d}{dt} \log(HC^2), \quad s = \frac{1}{H} \frac{d}{dt} \log(c_s), \quad (40)$$

as well as

$$\lambda' \equiv C^3(M_2^4 + M_3^4) + C^2(2\hat{M}_1^3 - \hat{M}_2^3) \left(3 - \frac{\mathcal{N}}{M_P^2}\right). \quad (41)$$

Furthermore, $\delta L/\delta\zeta$ denotes the variation of the quadratic action by ζ ,

$$\frac{\delta L}{\delta\zeta} = -2M_{\text{Pl}}^2 \partial_t(a\partial^2\chi) + 2a\mathcal{N}c_s^2\partial^2\zeta, \quad (42)$$

and the coefficient of this term in the action (39) is

$$g(\zeta) = \frac{n(HC)^2}{4c_s^2}\zeta^2 + \frac{HC^2}{c_s^2}\zeta\dot{\zeta} + \frac{C^2}{4a^2} [-(\partial_i\zeta)^2 + \partial^{-2}(\partial_i\partial_j(\partial_i\zeta\partial_j\zeta))] + \\ - \frac{C}{2a^2} [\partial_i\zeta\partial_i\chi - \partial^{-2}(\partial_i\partial_j(\partial_i\zeta\partial_j\chi))]. \quad (43)$$

As a quick consistency check, we note that our Eq.(39) reduces to an analogous expression, Eq. (4.26) of Ref. [16], with the following substitutions: $\hat{M}_1^3 = \hat{M}_2^3 = 0$, $C \rightarrow -H^{-1}$, $\mathcal{N}c_s^2 \rightarrow \varepsilon$, $n \rightarrow \eta \equiv (d\log\varepsilon/dt)/H$ and $\varrho \rightarrow \varepsilon$. In that case, a small speed of sound implies large (equilateral) non-Gaussianity, $f_{NL}^{\text{equil}} \sim 1/c_s^2$. Our result generalizes the cubic action of [16] to the case of non-zero \hat{M}_1^3 and \hat{M}_2^3 , which opens up qualitatively novel ways of generating large non-Gaussianity, as explained in the main text.

B Bispectrum

We follow the standard nomenclature in defining various quantities associated with the two- and three-point functions of the comoving curvature perturbation ζ , see, e.g., Ref. [31]. The three-point function is defined as

$$\langle\zeta(\vec{k}_1)\zeta(\vec{k}_2)\zeta(\vec{k}_3)\rangle = (2\pi)^3\delta^{(3)}\left(\sum_i\vec{k}_i\right) B_\zeta(k_1, k_2, k_3), \quad (44)$$

and $k_t \equiv k_1 + k_2 + k_3$. We will at times use a related quantity, $B_\Phi = (3/5)^3 B_\zeta$, for the Bardeen potential $\Phi = (3/5)\zeta$. Moreover, we will denote

$$A = H^2\mathcal{N}, \quad (45)$$

where \mathcal{N} is the normalization factor for the ζ kinetic term. The scalar perturbations' power spectrum can be directly read off Eq. (34)

$$P_\zeta(k) = \frac{1}{4} \frac{H^4}{A} \frac{1}{(c_s k)^3}, \quad (46)$$

and dimensionless power spectrum is

$$\Delta_\zeta^2 = \frac{k^3}{2\pi^2} P_\zeta(k) \simeq 5\pi^2 \cdot 10^{-9} . \quad (47)$$

With all of the above qualifications, the contribution from each of the relevant cubic operators (we only exclude the operators which are suppressed by at least two powers of slow-roll) in (39) to the bispectrum reads

$$B_\Phi(k_1, k_2, k_3) = 2 \left(\frac{3}{20} \right)^3 \frac{H^{10}}{A^3 c_s^6} \sum_i c_i S_i(k_1, k_2, k_3) + \text{cyclic} . \quad (48)$$

1. Contribution from $\dot{\zeta}^3$

$$S_1(k_1, k_2, k_3) = \frac{4}{k_t^3 k_1 k_2 k_3} \quad (49)$$

$$c_1 = -\frac{M_{\text{Pl}}^2}{c_s^4} \frac{\alpha^2 - \alpha - \varepsilon + c_s^2(3\alpha - \gamma + 1)(\alpha^2 - \alpha - \varepsilon) + c_s^4(\alpha - 1)(6\alpha^2 - 3\alpha\gamma + 3\gamma + \delta - 2\varepsilon)}{(\alpha - 1)^4} \quad (50)$$

2. Contribution from $\zeta \dot{\zeta}^2$

$$S_2(k_1, k_2, k_3) = 2 \frac{k_t + k_1}{k_t^2 k_1^3 k_2 k_3} \quad (51)$$

$$c_2 = 3 \frac{M_{\text{Pl}}^2}{c_s^4} \frac{(\alpha^2 - \alpha - \varepsilon)(1 - (\alpha - 1)^2 c_s^2)}{(\alpha - 1)^4} \quad (52)$$

3. Contribution from $\zeta(\partial\zeta)^2$

$$S_3(k_1, k_2, k_3) = (k_1^2 - k_2^2 - k_3^2) \left(-\frac{k_t}{(k_1 k_2 k_3)^3} + \frac{\sum_{i>j} k_i k_j}{k_t (k_1 k_2 k_3)^3} + \frac{1}{k_t^2 (k_1 k_2 k_3)^2} \right) \quad (53)$$

$$c_3 = -\frac{c_2}{3} \quad (54)$$

4. Contribution from $\dot{\zeta} \partial_i \zeta \partial_i \partial^{-2} \dot{\zeta}$

$$S_4(k_1, k_2, k_3) = \frac{k_1^2 - k_2^2 - k_3^2}{2k_t k_1 (k_2 k_3)^3} \left(2 + \frac{k_2 + k_3}{k_t} \right) \quad (55)$$

$$c_4 = -\frac{M_{\text{Pl}}^2}{c_s^4} \frac{(\alpha^2 - 5\alpha - \varepsilon + 4)(\alpha^2 - \alpha - \varepsilon)^2}{2(\alpha - 1)^6} \quad (56)$$

5. Contribution from $\partial^2 \zeta \partial_i \partial^{-2} \dot{\zeta} \partial_i \partial^{-2} \dot{\zeta}$

$$S_5(k_1, k_2, k_3) = \frac{k_1^2 - k_2^2 - k_3^2}{k_t k_1 (k_2 k_3)^3} \left(1 + \frac{k_1}{k_t} \right) \quad (57)$$

$$c_5 = \frac{M_{\text{Pl}}^2}{c_s^4} \frac{(\alpha^2 - \alpha - \varepsilon)^2 (3\alpha^2 - 3\alpha + \varepsilon)}{4(\alpha - 1)^6} \quad (58)$$

6. Contribution from $\partial^2 \zeta \partial_i \zeta \partial_i \partial^{-2} \dot{\zeta}$

$$S_6(k_1, k_2, k_3) = \frac{k_1^2 - k_2^2 - k_3^2}{2k_t k_1 (k_2 k_3)^3} \left(2 + \frac{2k_1 + k_2 + k_3}{k_t} + \frac{2k_1(k_2 + k_3)}{k_t^2} \right) \quad (59)$$

$$c_8 = -\frac{M_{\text{Pl}}^2}{c_s^4} \frac{2\alpha(\alpha^2 - \alpha - \varepsilon)}{(\alpha - 1)^4} \quad (60)$$

7. Contribution from $\partial^2 \zeta (\partial \zeta)^2$

$$S_7(k_1, k_2, k_3) = 2 \frac{k_1^2 - k_2^2 - k_3^2}{k_t k_1 (k_2 k_3)^3} \left(1 + \frac{\sum_{i>j} k_i k_j}{k_t^2} + \frac{3k_1 k_2 k_3}{k_t^3} \right) \quad (61)$$

$$c_7 = \frac{M_{\text{Pl}}^2}{c_s^4} \frac{\alpha}{(\alpha - 1)^3} \quad (62)$$

8. Contribution from $\partial^2 \zeta \dot{\zeta}^2$

$$S_8(k_1, k_2, k_3) = 4 \frac{k_t + 3k_1}{k_t^4 k_1 k_2 k_3} \quad (63)$$

$$c_8 = \frac{M_{\text{Pl}}^2}{c_s^2} \frac{4\alpha - \gamma}{(\alpha - 1)^3} \quad (64)$$

We have left out the operator $\zeta^2 \dot{\zeta}$ in (39), since its coefficient is at most of order $\sim \varepsilon^2$, and it is not expected to affect our analysis in any significant way. Moreover, the last term in Eq. (39), being proportional to the lower-order ζ equation of motion, can be removed by a field redefinition. The latter redefinition also contributes to the three-point function of the conserved scalar mode through the function $g(\zeta)$ in (43). The contributions of the terms in this function that include derivatives are suppressed at superhorizon distances. Moreover, the first term contributes to f_{NL} by an amount that scales as η/c_s^2 (see, e.g. [16]). We neglect this piece in our analysis, since it is always expected to be sub-dominant (whenever non-trivial constraints from bispectrum arise, there are leading contributions, enhanced by at least a factor of $1/\eta$ compared to it.)

C The regime with $f_{\text{NL}} \propto \frac{1}{c_s^6}$

In this appendix, we show how the behavior (6) arises in the model (2.3). For our purposes, it will be sufficient to set $\varepsilon = 0$ to avoid over-complication of expressions. In order to study the part of the parameter space, described by (19), it will prove convenient to slightly change the notation by defining

$$x \equiv \frac{1 - \alpha}{c_s^2}, \quad (65)$$

so that x is generically an order-one constant. The normalization factor for the curvature perturbation then becomes

$$\mathcal{N} = M_{\text{Pl}}^2 \frac{1 - xc_s^2}{xc_s^4}. \quad (66)$$

The latter quantity has a strong dependence on the speed of sound: for small c_s^2 , it grows like $\mathcal{N} \propto 1/c_s^4$, and this appears to make the scalar perturbations weakly coupled – suppressing the self-interactions of the canonically normalized ζ , and therefore suppressing non-Gaussianity. This observation is decisive, however: in order to make a conclusive statement regarding non-Gaussianity, one has to study the c_s -dependence of the cubic ζ interactions in the theory at hand. In fact, we will find that the cubic interactions grow as fast as $1/c_s^{10}$, eventually leading to non-Gaussianity of order $f_{\text{NL}} \propto 1/c_s^6$. The action for the comoving curvature perturbation, up to the cubic order in non-linearity, is given in Eq. (39). In the remainder of this section, we will confine ourselves to the leading order in the $1/c_s^2$ -expansion, which allows to extract the fastest-growing effects in the deep subluminal region of the parameter space of interest. There are seven operators that contribute in this limit, and assuming $\alpha \simeq 1$, the relevant cubic action becomes

$$\begin{aligned} S_\zeta^{(3)} = & M_{\text{Pl}}^2 \int d^4x a^3 \left\{ \frac{1}{xc_s^4} \left[\dot{\zeta}^2 - c_s^2 \frac{(\partial\zeta)^2}{a^2} \right] + \frac{1}{x^3 c_s^{10}} \left[\frac{1}{H} \dot{\zeta}^3 - 3\zeta \dot{\zeta}^2 + c_s^2 \dot{\zeta} \frac{(\partial\zeta)^2}{a^2} \right. \right. \\ & \left. \left. - \frac{3}{2} \dot{\zeta} \partial_i \zeta \partial_i \partial^{-2} \dot{\zeta} - \frac{3}{4} \partial^2 \zeta \partial_i \partial^{-2} \dot{\zeta} \partial_i \partial^{-2} \dot{\zeta} + 2 \frac{c_s^2}{H} \frac{\partial^2 \zeta}{a^2} \partial_i \zeta \partial_i \partial^{-2} \dot{\zeta} - \frac{c_s^4}{H^2} \frac{\partial^2 \zeta}{a^2} \frac{(\partial\zeta)^2}{a^2} \right] \right\}. \quad (67) \end{aligned}$$

Various terms in (67) appear to be of different order in $1/c_s^2$; however, one should keep in mind that higher spatial derivatives lead to additional factors of $1/c_s$ in the amplitude of non-Gaussianity, so that e.g. the two cubic operators $\dot{\zeta}^3/c_s^{10}$ and $\zeta(\partial\zeta)^2/c_s^8$ contribute comparably to f_{NL} in the small c_s^2 limit.

Summing up all the contributions to non-Gaussianity described in the previous appendix yields a simple expression at the leading order in the $1/c_s^2$ expansion

$$B_\Phi(k_1, k_2, k_3) = -12 \left(\frac{3}{20} \right)^3 \frac{M_{\text{Pl}}^2 H^{10}}{(Ax)^3 c_s^{16}} \frac{1}{k_t^3 (k_1 k_2 k_3)^2} \left[\sum_i k_i^3 - \sum_{i \neq j} k_i^2 k_j + 2k_1 k_2 k_3 \right]. \quad (68)$$

The shape of the bispectrum is close to the equilateral one, see Fig. 6. We note, that the second and the third operators above lead to squeezed non-Gaussianity, which can not characterize a derivatively coupled theory of the sort we are considering; and indeed, all ‘non-equilateralness’ cancels out in the full bispectrum, which is a nice consistency check of our results. This leaves us with the simple expression in Eq. (68).

The fact that all of these terms are equally important at typical frequency scales of order of the inflationary Hubble rate can also be seen from the analysis of the $2 \rightarrow 2$ scattering of ζ quanta. It is a fairly generic fact, that the strong coupling scale of a theory shrinks to zero (or, more precisely, becomes dominated by higher-order effects in the dispersion relation) as the speed of sound is taken to vanish, and in the limit of small c_s^2 that we are interested in

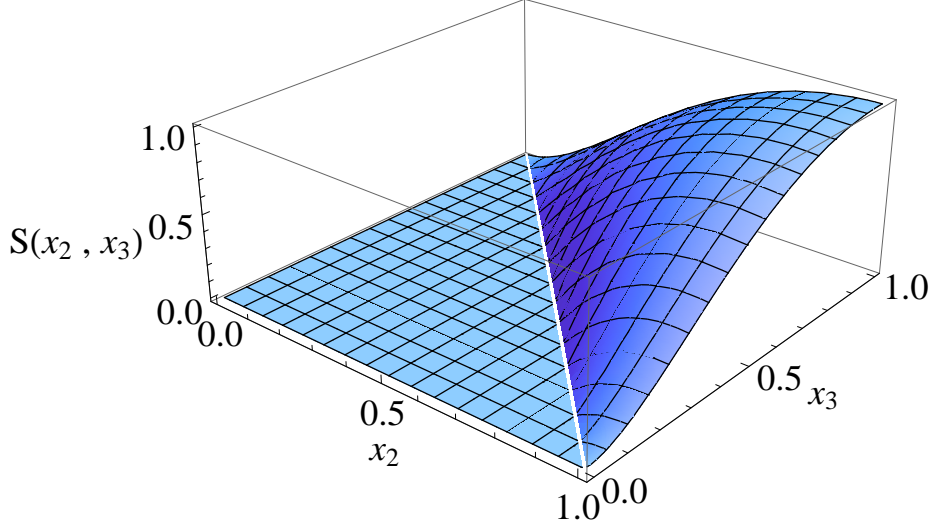


Figure 6: The shape of non-Gaussianity, corresponding to Eq. (68) (we use the standard notation: $x_2 = k_2/k_1$ and $x_3 = k_3/k_1$.)

here, one should be extra careful with this scale. In particular, in order to be able to trust our low-energy effective theory, we should make sure that the strong coupling scale of that theory is parametrically greater than the Hubble rate H — the typical energy/frequency scale that we measure the inflationary observables at. In order to assess the strong coupling scale Λ_\star associated with the i -th cubic interaction in (67), one can study a $2 \rightarrow 2$ scattering of ζ that stems from that interaction. These interactions are non-relativistic, however, and extra care has to be taken in order to properly define the energy and momentum scales at which perturbative unitarity is violated in the scattering of interest. There are various ways of addressing these issues, and perhaps the shortest and the most straightforward one is via formally switching back to the relativistic notation. To this end, one can rescale time variable in the action (67), so that $t = \tilde{t}/c_s$, and $\omega = \tilde{\omega}c_s$ for the frequency. The quadratic ζ action then becomes

$$S_\zeta^{(2)} = M_{\text{Pl}}^2 \int d^3x d\tilde{t} a^3 \frac{1}{xc_s^3} \left[(\partial_{\tilde{t}}\zeta)^2 - \frac{(\partial\zeta)^2}{a^2} \right], \quad (69)$$

while the first cubic interaction in (67) reads

$$S_\zeta^{(3)} \supset M_{\text{Pl}}^2 \int d^3x d\tilde{t} a^3 \left[\frac{1}{x^3 c_s^8 H} (\partial_{\tilde{t}}\zeta)^3 + \dots \right]. \quad (70)$$

Canonically normalizing the curvature perturbation, $\zeta = x^{1/2} c_s^{3/2} \zeta_c$, one immediately obtains the strong coupling scale associated with the given operator $\tilde{\Lambda}_\star^2 = x^{3/2} c_s^{7/2} M_{\text{Pl}} H$. Note that this scale is still defined in the rescaled temporal coordinates; going back to the original coordinates, one finally arrives at the true strong coupling (energy) scale

$$\Lambda_\star^2 = c_s^2 \tilde{\Lambda}_\star^2 = x^{3/2} c_s^{11/2} M_{\text{Pl}} H. \quad (71)$$

Demanding that this scale be larger than the inflationary Hubble rate yields the following bound

$$x^{3/2} c_s^{11/2} M_{\text{Pl}} > H . \quad (72)$$

What about the rest of the cubic operators in (67)? We have seen, that they all contribute by an equal order of magnitude to non-Gaussianity, so that there should exist a well-defined sense in which they are all ‘equally strongly coupled’ around the Hubble frequencies. Repeating the above analysis, it is easy to see that the six remaining cubic interactions in (67) imply strong coupling scales, in general different from Λ_\star . For example, the operator $\zeta \dot{\zeta}^2$ starts violating perturbative unitarity in a $2 \rightarrow 2$ scattering of ζ quanta around the frequency scales of order

$$\bar{\Lambda}_\star = x^{3/2} c_s^{11/2} M_{\text{Pl}} . \quad (73)$$

Being different from Λ_\star , the expression for $\bar{\Lambda}_\star$ nevertheless implies the exact same condition (72) if the scattering at Hubble frequencies is to be unitary. One can straightforwardly check that the same conclusion applies to all operators in (72), fixing the sense in which all of these operators are equally important for the physics at the horizon.

References

- [1] P. Creminelli, M. A. Luty, A. Nicolis, and L. Senatore, “Starting the Universe: Stable Violation of the Null Energy Condition and Non-standard Cosmologies,” *JHEP* **0612** (2006) 080, [arXiv:hep-th/0606090](#) [[hep-th](#)].
- [2] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, “The Effective Field Theory of Inflation,” *JHEP* **0803** (2008) 014, [arXiv:0709.0293](#) [[hep-th](#)].
- [3] S. Weinberg, “Effective Field Theory for Inflation,” *Phys. Rev.* **D77** (2008) 123541, [arXiv:0804.4291](#) [[hep-th](#)].
- [4] C. Armendariz-Picon, T. Damour, and V. F. Mukhanov, “k - inflation,” *Phys.Lett.* **B458** (1999) 209–218, [arXiv:hep-th/9904075](#) [[hep-th](#)].
- [5] E. Silverstein and D. Tong, “Scalar speed limits and cosmology: Acceleration from D-cceleration,” *Phys.Rev.* **D70** (2004) 103505, [arXiv:hep-th/0310221](#) [[hep-th](#)].
- [6] M. Alishahiha, E. Silverstein, and D. Tong, “DBI in the sky,” *Phys.Rev.* **D70** (2004) 123505, [arXiv:hep-th/0404084](#) [[hep-th](#)].
- [7] L. Senatore, K. M. Smith, and M. Zaldarriaga, “Non-Gaussianities in Single Field Inflation and their Optimal Limits from the WMAP 5-year Data,” *JCAP* **1001** (2010) 028, [arXiv:0905.3746](#) [[astro-ph.CO](#)].
- [8] G. D’Amico and M. Kleban, “Non-Gaussianity after BICEP2,” *Phys. Rev. Lett.* **113** (2014) 081301, [arXiv:1404.6478](#) [[astro-ph.CO](#)].

- [9] D. Pirtskhalava, L. Santoni, E. Trincherini, and F. Vernizzi, “Weakly Broken Galileon Symmetry,” [arXiv:1505.00007 \[hep-th\]](#).
- [10] A. Nicolis, R. Rattazzi, and E. Trincherini, “The Galileon as a local modification of gravity,” *Phys.Rev.* **D79** (2009) 064036, [arXiv:0811.2197 \[hep-th\]](#).
- [11] B. W. Lee, C. Quigg, and H. Thacker, “Weak Interactions at Very High-Energies: The Role of the Higgs Boson Mass,” *Phys.Rev.* **D16** (1977) 1519.
- [12] S. Mizuno and K. Koyama, “Primordial non-Gaussianity from the DBI Galileons,” *Phys.Rev.* **D82** (2010) 103518, [arXiv:1009.0677 \[hep-th\]](#).
- [13] C. Burrage, C. de Rham, D. Seery, and A. J. Tolley, “Galileon inflation,” *JCAP* **1101** (2011) 014, [arXiv:1009.2497 \[hep-th\]](#).
- [14] D. Pirtskhalava, L. Santoni, E. Trincherini, and F. Vernizzi, “Large Non-Gaussianity in Slow-Roll Inflation,” [arXiv:1506.06750 \[hep-th\]](#).
- [15] J. M. Maldacena, “Non-Gaussian features of primordial fluctuations in single field inflationary models,” *JHEP* **0305** (2003) 013, [arXiv:astro-ph/0210603 \[astro-ph\]](#).
- [16] X. Chen, M.-x. Huang, S. Kachru, and G. Shiu, “Observational signatures and non-Gaussianities of general single field inflation,” *JCAP* **0701** (2007) 002, [arXiv:hep-th/0605045 \[hep-th\]](#).
- [17] T. Kobayashi, M. Yamaguchi, and J. Yokoyama, “Primordial non-Gaussianity from G-inflation,” *Phys.Rev.* **D83** (2011) 103524, [arXiv:1103.1740 \[hep-th\]](#).
- [18] P. Creminelli, J. Gleyzes, J. Norena, and F. Vernizzi, “Resilience of the standard predictions for primordial tensor modes,” [arXiv:1407.8439 \[astro-ph.CO\]](#).
- [19] **Planck** Collaboration, P. A. R. Ade *et al.*, “Planck 2015 results. XX. Constraints on inflation,” [arXiv:1502.02114 \[astro-ph.CO\]](#).
- [20] T. Kobayashi, M. Yamaguchi, and J. Yokoyama, “G-inflation: Inflation driven by the Galileon field,” *Phys.Rev.Lett.* **105** (2010) 231302, [arXiv:1008.0603 \[hep-th\]](#).
- [21] C. Deffayet, O. Pujolas, I. Sawicki, and A. Vikman, “Imperfect Dark Energy from Kinetic Gravity Braiding,” *JCAP* **1010** (2010) 026, [arXiv:1008.0048 \[hep-th\]](#).
- [22] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, “Causality, analyticity and an IR obstruction to UV completion,” *JHEP* **10** (2006) 014, [arXiv:hep-th/0602178 \[hep-th\]](#).
- [23] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, “Limits on non-gaussianities from wmap data,” *JCAP* **0605** (2006) 004, [arXiv:astro-ph/0509029 \[astro-ph\]](#).

- [24] A. A. Starobinsky, “A New Type of Isotropic Cosmological Models Without Singularity,” *Phys. Lett.* **B91** (1980) 99–102.
- [25] X. Chen, “Running non-Gaussianities in DBI inflation,” *Phys. Rev.* **D72** (2005) 123518, [arXiv:astro-ph/0507053](#) [astro-ph].
- [26] **Planck** Collaboration, P. Ade *et al.*, “Planck 2015 results. XVII. Constraints on primordial non-Gaussianity,” [arXiv:1502.01592](#) [astro-ph.CO].
- [27] **BICEP2 s** Collaboration, K. Array *et al.*, “BICEP2 / Keck Array VI: Improved Constraints On Cosmology and Foregrounds When Adding 95 GHz Data From Keck Array,” [arXiv:1510.09217](#) [astro-ph.CO].
- [28] N. Arkani-Hamed, P. Creminelli, S. Mukohyama, and M. Zaldarriaga, “Ghost inflation,” *JCAP* **0404** (2004) 001, [arXiv:hep-th/0312100](#) [hep-th].
- [29] L. Senatore, “Tilted ghost inflation,” *Phys.Rev.* **D71** (2005) 043512, [arXiv:astro-ph/0406187](#) [astro-ph].
- [30] M. M. Ivanov and S. Sibiryakov, “UV-extending Ghost Inflation,” *JCAP* **1405** (2014) 045, [arXiv:1402.4964](#) [astro-ph.CO].
- [31] D. Babich, P. Creminelli, and M. Zaldarriaga, “The Shape of non-Gaussianities,” *JCAP* **0408** (2004) 009, [arXiv:astro-ph/0405356](#) [astro-ph].